Short Communication

Resolving the database sequence discrepancies for the *Staphylococcus aureus* bacteriophage φ11 amidase

David M. Donovan, Juli Foster-Frey, Wesley M. Garrett and LeAnn Blomberg

Animal Biosciences and Biotechnology Laboratory, Animal and Natural Resources Institute, BARC, ARS, USDA, Baltimore Ave, Beltsville, MD, USA

There are two conflicting primary nucleotide sequences of the *Staphylococcus aureus* bacteriophage φ11 amidase gene in Genbank. Nucleotide sequence differences as well as alternative translational start site assignments result in three non-identical protein sequence predictions for this amidase. Therefore, it is prudent to verify the correct phi11 amidase protein sequence, especially since multiple versions of the amidase gene have been subcloned, deletion analysis performed, and their experimental use described. There is also a resurgence of interest in the expression and use of bacteriophage lytic proteins as bactericidal agents and the φ11 amidase has a high antimicrobial potential. The correct amidase sequence is identified through a combination of DNA sequence analysis and matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry analysis of the recombinant purified phi11 amidase protein.

Keywords: Peptidoglycan hydrolase / Phage 80α / Endolysin / Lysin

Received: June 05, 2007; accepted July 18, 2007

DOI 10.1002/jobm.200700179

Introduction

In this age of ever-increasing antimicrobial resistance among staphylococci, novel antimicrobials with activity against multi-drug resistant pathogens are in high demand for clinical applications world-wide. Peptidoglycan hydrolases (endolysins) encoded by bacteriophage naturally digest cell wall peptidoglycan bonds in order to allow nascent phage particles to escape infected cells. Many phage lysins are good candidate antimicrobials, able to lyse their target pathogen when exposed externally. They also show high species-specificity allowing them to avoid a major pitfall of broad range antimicrobials (e.g. penicillin, tetracycline), by reducing the likelihood of resistance development in non-targeted, commensal strains [1].

The phi11 amidase is a peptidoglycan hydrolase with multiple domains. The predicted amino acid (AA) sequence of the phi11 amidase protein indicates two lytic domains and a C-terminal SH3b cell wall binding domain [2, 3]. The N-terminal lytic domains consist of a cysteine, histidine-dependent amidohydrolases/peptidase (CHAP) endopeptidase [4] and an amidase domain [5] that can both degrade staphylococcal peptidoglycan [3, 6]. Purified recombinant phi11 amidase also lyses viable [7] or heat/SDS treated staphylococci [8]. Deletion analysis has demonstrated that the N-terminal CHAP endopeptidase domain [7, 8] does not require the C-terminal SH3b domain for full activity while the mid-protein amidase domain is dependent on the SH3b domain for activity [8]. Viable *Staphylococcus aureus,* including those in biofilms, and multiple coagulase negative staphylococci are susceptible to the lytic action of the phi11 amidase, making it a strong candidate antimicrobial [7, 8].

Materials and methods

There are two distinct DNA sequences and three predicted protein sequences in GenBank for the bacteriophage phi11 amidase gene (also called *S. aureus* NCTC 8325 autolysin) (Table 1). The first report in 1990 describes a *S. aureus* PS47 genomic fragment encoding LytA, a *S. aureus* autolysin [9]. Further analysis by the

Correspondence: David M. Donovan, Animal Biosciences and Biotechnology Laboratory, Animal and Natural Resources Institute, BARC, ARS, USDA, 10300 Baltimore Ave, Beltsville, MD 20705-2350
E-mail: ddonovan@anri.barc.usda.gov
Tel.: 301-504-9150
Fax: 301-504-8571

Table 1. Genbank Accession numbers for protein and nucleotide sequences of the phi11 amidase/*S. aureus* NCTC 8325 autolysin.

	protein Acc. #	nucleotide Acc. #
phi11LytA	P24556	M76714
phi11ORF53	NP_803306.1	NC_004615
autolysin	YP_500516	NC_007795

same group identified the protein as the phi11 prophage amidase [10]. The second GenBank submission was derived from the phi11 prophage genomic sequence. The prophage genomic sequence was extracted from the *S. aureus* 8325 genomic sequence analysis and was reported independently in a study comparing phi11 to the phi12 and phi13 genomes [11]. The phi11ORF53 from this work was PCR amplified and described as the phi11 amidase in later works by Sass and Bierbaum [8]. The Iandolo group has more recently submitted to GenBank (Table 1) a predicted autolysin protein with an identical nucleotide sequence as phi11ORF53 that was derived from a description of the *S. aureus* NCTC 8325 genome [12].

Results and discussion

Due to the high interest in bacteriophage proteins as antimicrobials [13, 14] and the phi11 amidase in particular, it is important to define the exact DNA and protein sequences being studied. This becomes especially important when interpreting deletion analyses used to study the phi11 amidase domains [7, 8]. Both phi11ORF53 and phi11LytA versions of the phi11 amidase have been introduced into the pET21a expression vector containing a C-terminal His-tag that enables nickel chromatography purification [6–8]. These plasmid subcloning events utilize high-fidelity PCR to amplify the region of the nucleotide sequence containing the open reading frame (ORF) of the predicted amidase protein. By introducing this ORF into the pET21a vector, the native phage promoter elements are replaced with high level expression promoters engineered into these inducible plasmids. Due to the use of engineered promoter elements and predicted ORFs in these expression constructs, accurate ORF assignment is essential.

A BLAST search of Genbank with the phi11ORF53 amidase protein sequence identifies a family of over 15 staphylococcal protein entries with >90% identity at the level of predicted amino acid sequence (data not shown). The phi11ORF53 primary nucleotide sequence (identical to the *S. aureus* autolysin sequence) is identical to phi11LytA sequence with the exception of three bases, which result in two regions of predicted amino acid sequence divergence (Fig. 1). First, an A to T nucleotide (nt) substitution in phi11LytA at nt 326 results in an AA substitution of an Arg to Trp in the phi11LytA predicted amino acid sequence at residue 13. Second, a C is deleted in the phi11LytA sequence at nt 666 that creates a frame shift in the predicted amino acid sequence beginning at phi11LytA AA 127 (Fig. 1). The frame shift is corrected by the insertion of a C in the phi11LytA sequence at nt 788, bringing the AA sequence back in frame with phi11ORF53 for the C-terminal region of the protein. The nucleotide alterations result in a stretch of 40 non-conserved AA spanning AA 127 to AA 166 that are dramatically different compared to phi11ORF53 (Fig. 1) and the other family members (data not shown). The C-terminal half of all three versions of the phi11 amidase (and autolysin) gene are perfectly conserved.

A protein BLAST search against the nr (non-redundant) sequences identified 15 predicted protein sequence entries in GenBank with a homology >90% to the phi11ORF53 and autolysin versions (data not shown). The nearly identical 15 predicted protein sequences lack the first nine N-terminal amino acids assigned to phi11ORF53 but share near perfect identity with the 40 amino acid mid-protein region of the predicted phi11ORF53. The phi11LytA and phage 80α amidase (Genbank #AAB39699) both have the identical 40 AA mid-protein sequence that differs significantly from the other family members. The finding of two non-conserved members among a family of proteins with such a high level of conservation begs the question of whether or not these protein differences represent divergence within this highly conserved family of proteins or are the result of errors in the sequence analysis. Sequence discrepancies resulting from potential sequencing errors are not uncommon in public databases [15].

To address the phi11ORF53 nine additional N-terminal amino acid residues, the nucleotide sequence of the phi11ORF53 versus phi11LytA were compared in the region spanning the promoter and translation start site (Fig. 1A). The nucleotide sequences in this region are identical between these GenBank entries. The phi11LytA submission to GenBank describes the location of the consensus prokaryotic transcriptional promoter and translation expression signals: –35 box, –10 box and ribosome binding site [10, 16]. However, the identification of these three regulatory elements were not mentioned in either the study comparing the phi11 genome to that of phi12 and phi13 where phi11ORF53 was first identified, [11] or the subsequent GenBank

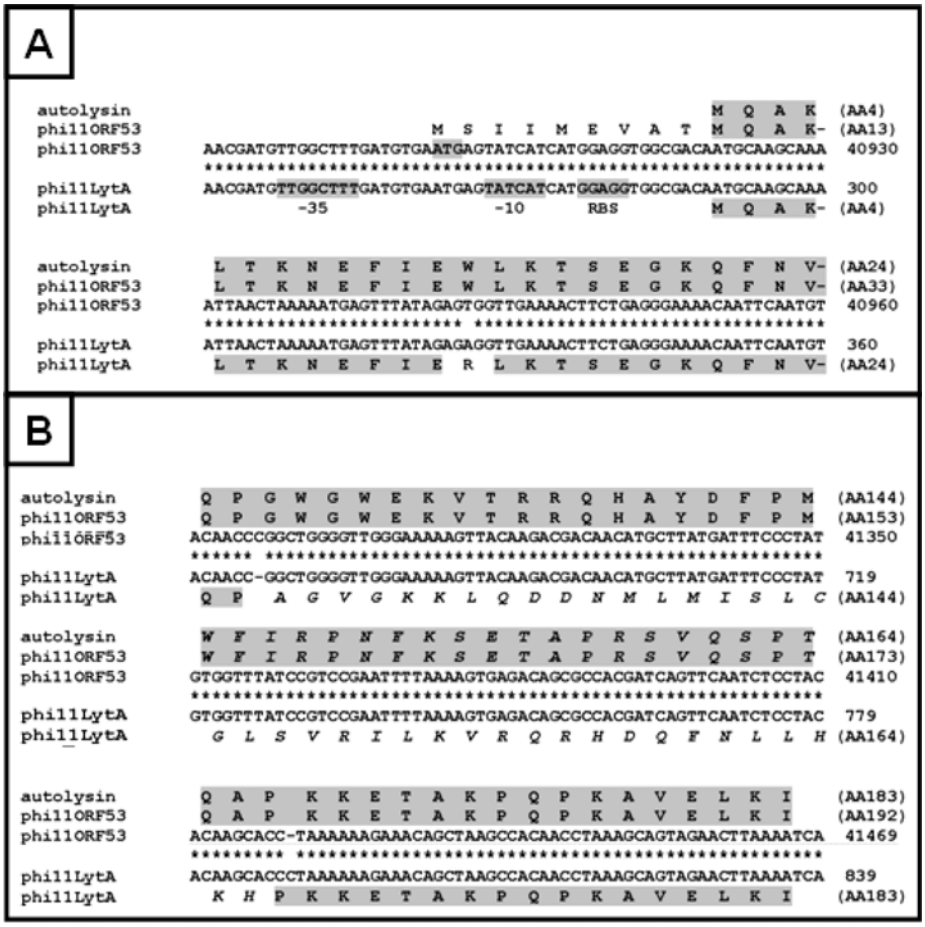
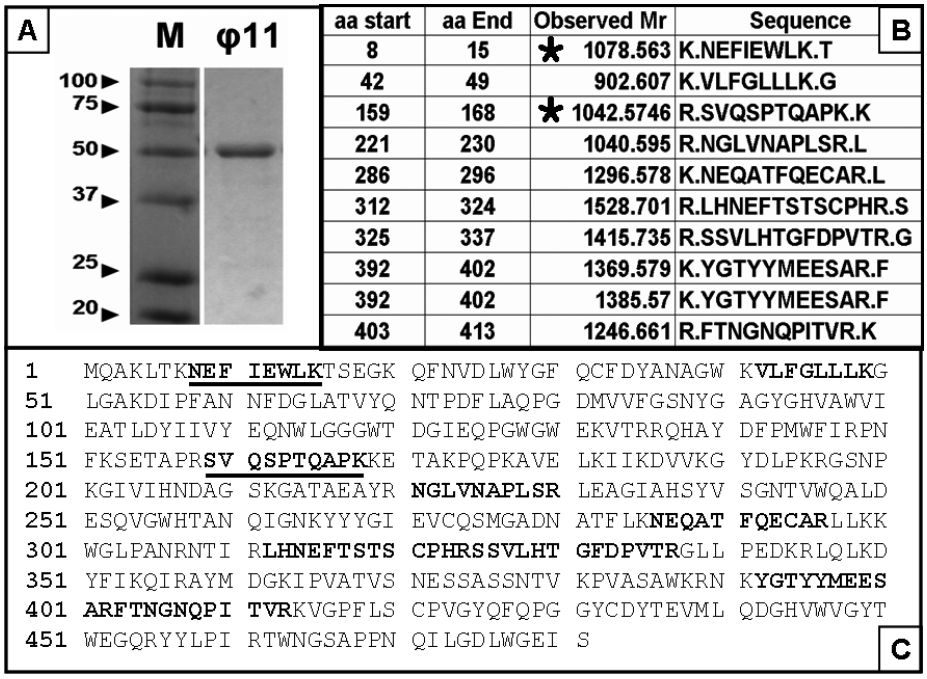


Figure 1. Nucleotide sequence alignment and predicted amino acid sequence of phi11 amidase homologues. A. Sequence alignment in the promoter and N-terminal region. Shaded regions of the nucleotide sequence indicate conserved prokaryotic promoter elements. B. Sequence alignment in the mid-protein 40 amino acid non-identical region. Shaded regions of the protein sequence indicate AA sequences >90% conserved with 15 other Genbank entries (data not shown). Asterisks indicate identity between the sequences. The phi11 prophage ORF53 genomic sequence and the autolysin nucleotide sequences are identical, having been derived from the same *S. aureus* 8325 genomic nucleotide sequence, with the difference between these submissions being an altered assignment of the translational start site methionine codon. (AA#) indicates amino acid number.

submission (NC_004615), or where the same sequence is referred to as an *S. aureus* NCTC8325 autolysin [12] or its Genbank submission (YP_500516). Upon examination of the prophage ORF53 genomic sequence, consensus prokaryotic transcriptional and translational elements (–35 box, –10 box and ribosome binding site) are lacking in the putative promoter region, 5′ to the ORF53 translational initiation methionine codon. The more traditional translation start site identified for phi11LytA (located 10 codons 3′ to the ORF53 translational start site) was adopted by the Iandolo group later in 2006 with the *S. aureus* NCTC8325 autolysin (Genbank submission YP_500516). Despite this N-terminal discrepancy, the N-terminal nine amino acids of ORF53 do not appear detrimental to the function of the recombinant phi11 protein. Sass and Bierbaum have shown strong lytic activity of both the full length ORF53 protein, and a C-terminal deletion construct harboring just the N-terminal 180 AA. Both constructs include the ORF53 N-terminal 9 AA [8].

To identify which of the other two predicted amino acid sequences is correct, DNA sequence analysis and Peptide Mass Fingerprinting of phi11LytA by matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) mass spectrometry analysis were performed. DNA sequence analysis predicted the phi11ORF53 nucleotide sequence to be correct (data not shown). To provide a second line of evidence to support this finding, the full length His-tagged phi11LytA recombinant protein [7] derived from the phi11LytA genomic fragment described previously [9] was purified by nickel chromatography and resolved by denatur-



aa start	aa End	Observed Mr	Sequence
8	15	* 1078.563	K.NEFIEWLK.T
42	49	902.607	K.VLFGLLLK.G
159	168	* 1042.5746	R.SVQSPTQAPK.K
221	230	1040.595	R.NGLVNAPLSR.L
286	296	1296.578	K.NEQATFQECAR.L
312	324	1528.701	R.LHNEFTSTSCPHR.S
325	337	1415.735	R.SSVLHTGFDPVTR.G
392	402	1369.579	K.YGTYYMEESAR.F
392	402	1385.57	K.YGTYYMEESAR.F
403	413	1246.661	R.FTNGNQPITVR.K

```
1   MQAKLTKNEF IEWLKTSEGK QFNVDLWYGF QCFDYANAGW KVLFGLLLKG
51   LGAKDIPFAN NFDGLATVYQ NTPDFLAQPG DMVVFGSNYG AGYGHVAWVI
101 EATLDYIIVY EQNWLGGGWT DGIEQPGWGW EKVTRRQHAY DFPMWFIRPN
151 FKSETAPRSV QSPTQAPKKE TAKPQPKAVE LKIIKDVVKG YDLPKRGSNP
201 KGIVIHNDAG SKGATAEAYR NGLVNAPLSR LEAGIAHSYV SGNTVWQALD
251 ESQVGWHTAN QIGNKYYYGI EVCQSMGADN ATFLKNEQAT FQECARLLKK
301 WGLPANRNTI RLHNEFTSTS CPHRSSVLHT GFDPVTRGLL PEDKRLQLKD
351 YFIKQIRAYM DGKIPVATVS NESSASSNTV KPVASAWKRN KYGTYYMEES
401 ARFTNGNQPI TVRKVGPFLS CPVGYQFQPG GYCDYTEVML QDGHVWVGYT
451 WEGQRYYLPI RTWNGSAPPN QILGDLWGEI S
```

Figure 2. phi11LytA protein and sequence analysis by MALDI-TOF. A. SDS PAGE manufacturer's analysis of phi11LytA recombinant protein obtained from nickel chromatography purification. Ni-NTA purification was per suggested conditions (Qiagen, Valencia CA) as described previously [7] with these modifications: 2 ml Ni-NTA columns with protein bound were washed with various volumes of Wash Buffer (50mM Sodium Phosphate, 300 mM NaCl, pH 8.0) containing increasing concentrations of imidazole [10 ml Wash buffer 1 (10 mM imidazole); 20 ml Wash Buffer 2 (20 mM imidazole); eluted with 1.2 ml Wash Buffer 3 (250 mM imidazole)]. Lane M, Molecular Weight Markers (MW in kD); Lane φ11, 4 ug phillLytA protein (MW = 53,899). B. Tryptic peptide fragments identified by MALDI-TOF mass spectrometry. Asterisks denote sequences underlined in 2C, that clarify a region of discrepancy between phi11ORF53 and phi11LytA. C. phi11 amidase amino acid sequence. Highlighted sequences represent tryptic fragments that were identified by the Mascot search of the MALDI-TOF peptide mass fingerprints. Underlined fragments are those with Asterisks in B, and that clarify a region of discrepancy between phi11ORF53 and phi11LytA.

ing SDS PAGE (Fig. 2A). Following in-gel trypsin digestion [17], MALDI-TOF mass spectrometry analysis identified nine tryptic peptide fragments with molecular weights that matched predicted tryptic fragment molecular weights of the phi11 amidase. These fragments were identified as phi11 amidase fragments with the Mascot search engine (www.matrixscinece.com) (Fig. 2B). Two of the tryptic peptide fragments identified were within regions of the predicted phi11LytA protein sequence that contained discrepancies between the Genbank entries. Both tryptic fragments support the ORF53 (or autolysin) sequence in these regions. The first highlighted tryptic fragment, "NEFIEWLK", underlined in Fig. 2C demonstrates that phi11LytA contains a Trp rather than an Arg at AA 13. The second highlighted peptide sequence, "SVQSPTQAPK", underlined in Fig. 2C spans AA 159 to AA 168 and is within the region of the 40 AA mid-protein discrepancy (Fig. 1). Although correct sized fragments were predicted, there were no tryptic fragments identified that would support the divergent mid-protein 40 amino acid sequence of phi11LytA (Fig. 1). Thus it was concluded that the phi11LytA 40 amino acid mid-protein divergence does not exist in the recombinant protein derived from the phi11LytA subclone.

Due to an identical phi11LytA mid-protein sequence prediction for the phage 80α amidase entry obtained from a prophage genomic fragment (#AAB39699), we initially proposed an identical MALDI-TOF experiment to determine the mid-protein sequence for the phage 80α protein. However, the phage 80α genome was recently sequenced from phage DNA (as opposed to the prior prophage genomic fragment) and submitted to GenBank (#DQ517338). In this submission, the amidase protein (#ABF71642.1) does not show high homology to the phi11 amidase, so these experiments were no longer considered.

The *S. aureus* NCTC 8325 autolysin sequence (YP_500516) predicts the correct phi11 amidase protein sequence. The prokaryotic transcriptional and translational signals identified by Wang *et al.*, for phi11LytA define a conventional translational start site [10]. Redundant DNA sequence analysis of the phi11LytA ORF and MALDI-TOF analysis of phi11LytA recombinant

protein indicate that the ORF first defined by Iandola and colleagues [11] define the correct 40 amino acid mid-protein sequence of the phi11 amidase and that amino acid 13 (of the phi11LytA sequence) is Trp, not Arg. The lack of a 40 amino acid divergence between phi11LytA and phi11ORF53 is consistent with the similar results obtained from deletion analysis of the phi11 amidase gene in two independent labs [7, 8]. Collectively, these findings support the use of multiple diagnostic tools, such as DNA sequence regulatory element identification (e.g. promoter elements), multi-lab repeats of nucleotide sequence analysis and the use of protein sequence analysis, to correctly predict gene sequences and their translation products.

Acknowledgements

Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

References

[1] Loessner, M.J., 2005. Bacteriophage endolysins – current state of research and applications. Curr. Opin. Microbiol., **8**, 480–487.

[2] Ponting, C.P., Aravind, L., Schultz, J., Bork, P. and Koonin, E.V., 1999. Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. J. Mol. Biol., **289**, 729–745.

[3] Navarre, W.W., Ton-That, H., Faull, K.F. and Schneewind, O., 1999. Multiple enzymatic activities of the murein hydrolase from staphylococcal phage phi11. Identification of a D-alanyl-glycine endopeptidase activity. J. Biol. Chem., **274**, 15847–15856.

[4] Bateman, A. and Rawlings, N.D., 2003. The CHAP domain: a large family of amidases including GSP amidase and peptidoglycan hydrolases. Trends Biochem. Sci., **28**, 234–237.

[5] Rigden, D.J., Jedrzejas, M.J. and Galperin, M.Y., 2003. Amidase domains from bacterial and phage autolysins define a family of gamma-D,L-glutamate-specific amidohydrolases. Trends Biochem. Sci., **28**, 230–234.

[6] Grundling, A., Missiakas, D.M. and Schneewind, O., 2006. *Staphylococcus aureus* mutants with increased lysostaphin resistance. J. Bacteriol., **188**, 6286–6297.

[7] Donovan, D.M., Lardeo, M. and Foster-Frey, J., 2006. Lysis of staphylococcal mastitis pathogens by bacteriophage phi11 endolysin. FEMS Microbiol. Lett., **265**, 133–139.

[8] Sass, P. and Bierbaum, G., 2007. Lytic activity of recombinant bacteriophage {phi}11 and {phi}12 endolysins on whole cells and biofilms of *Staphylococcus aureus*. Appl. Environ. Microbiol., **73**, 347–352.

[9] Jayaswal, R.K., Lee, Y.I. and Wilkinson, B.J., 1990. Cloning and expression of a *Staphylococcus aureus* gene encoding a peptidoglycan hydrolase activity. J. Bacteriol., **172**, 5783–5788.

[10] Wang, X., Mani, N., Pattee, P.A., Wilkinson, B.J. and Jayaswal, R.K., 1992. Analysis of a peptidoglycan hydrolase gene from *Staphylococcus aureus* NCTC8325. J. Bacteriol., **174**, 6303–6306.

[11] Iandolo, J.J., Worrell, V., Groicher, K.H., Qian, Y., Tian, R., Kenton, S., Dorman, A., Ji, H., Lin, S., Loh, P., Qi, S., Zhu, H. and Roe, B.A., 2002. Comparative analysis of the genomes of the temperate bacteriophages phi11, phi12 and phi13 of *Staphylococcus aureus* 8325. Gene, **289**, 109–118.

[12] Gillaspy, A.F., Worrell, V., Orvis, J., Roe, B.A., Dyer, D.W. and Iandolo, J. J., 2006. The Staphylococcus aureus NCTC8325 Genome, In: Gram Positive Pathogens (V.A. Fischetti, R. Novick, J. Ferretti, D. Portnoy, J. Rood, eds.), pp. 381–412. ASM.

[13] Donovan, D.M., 2007. Bacteriophage and peptidoglycan degrading enzymes with antimicrobial applications. Recent Patents in Biotechnology, In press.

[14] Kwan, T., Liu, J., Dubow, M., Gros, P. and Pelletier, J., 2005. The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. Proc. Natl. Acad. Sci. USA, **102**, 5174–5179.

[15] Wesche, P.L., Gaffney, D.J. and Keightley, P.D., 2004. DNA sequence error rates in Genbank records estimated using the mouse genome as a reference. DNA Seq., **15**, 362–364.

[16] Busby, S. and Ebright, R.H., 1994. Promoter structure, promoter recognition, and transcription activation in prokaryotes. Cell, **79**, 743–746.

[17] Talbot, N.C., Blomberg, L.A., Mahmood, A., Caperna, T.J. and Garrett, W.M., 2007. Isolation and characterization of porcine visceral endoderm cell lines from *in vivo* 11-day blastocysts. In Vitro Cell. Dev. Biol., **43**, 72–86.

